

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

J. Tajima
4/2/01
Q63903
10f1
J1017 U.S. PTO
09/822513
04/02/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出 願 年 月 日
Date of Application:

2000年 3月31日

出 願 番 号
Application Number:

特願2000-099627

出 願 人
Applicant(s):

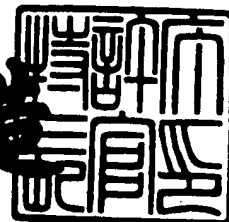
日本電気株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年12月 1日

特許庁長官
Commissioner,
Patent Office

及川耕造



【書類名】 特許願

【整理番号】 33509712

【提出日】 平成12年 3月31日

【あて先】 特許庁長官殿

【国際特許分類】 G06K 9/00
G06F 17/30

【発明者】

【住所又は居所】 東京都港区芝五丁目7番1号 日本電気株式会社内

【氏名】 田島 譲二

【特許出願人】

【識別番号】 000004237

【氏名又は名称】 日本電気株式会社

【代理人】

【識別番号】 100065385

【弁理士】

【氏名又は名称】 山下 穰平

【電話番号】 03-3431-1831

【手数料の表示】

【予納台帳番号】 010700

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9001713

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 映像録画方法およびそのシステムとその記録媒体

【特許請求の範囲】

【請求項 1】 複数フレームの映像信号中から入力顔画像を検出し、該入力顔画像と指定された問合せ顔画像とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号を録画することを特徴とする映像録画方法。

【請求項 2】 請求項 1 に記載の映像録画方法において、前記入力顔画像は標準顔画像により正規化して、前記問合せ顔画像を正規化した正規化顔画像とを照合し、所定の類似度以上に類似していた場合を前記照合に成功した場合とすることを特徴とする映像録画方法。

【請求項 3】 複数フレームの映像信号中から入力顔画像を検出し、該顔入力画像と指定された問合せ顔画像とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号の録画を行わないことを特徴とする映像録画方法。

【請求項 4】 請求項 3 に記載の映像録画方法において、前記入力顔画像は標準顔画像により正規化して、前記問合せ顔画像を正規化した正規化顔画像とを照合し、所定の類似度以上に類似していた場合を前記照合に成功した場合とすることを特徴とする映像録画方法。

【請求項 5】 対象物を検出して前記対象物が存在するフレームの前後の複数フレームを録画する映像録画システムにおいて、

映像信号から対象物の特徴点を検出する対象物検出手段と、該対象物の特徴点を利用して該対象物を正規化する対象物正規化手段と、予め問合せ画像を指定して正規化し正規化問合せ画像として出力する問合せ画像指定手段と、前記対象物正規化手段の出力と前記正規化問合せ画像とを照合する対象物照合手段と、前記対象物照合手段の結果に応じて前記映像信号を録画制御する制御手段とを備えたことを特徴とする映像録画システム。

【請求項 6】 請求項 5 に記載された映像録画システムにおいて、前記対象物検出手段は、前記映像信号の解像度を変換する解像度変換手段と、解像度変換

された映像信号から当該映像信号の特徴点を検出するフィルタリング手段と、前記映像信号の特徴点から、該特徴点の位置関係を検定して照合範囲信号を出力する仮説検定手段とからなることを特徴とする映像録画システム。

【請求項7】 請求項5に記載された映像録画システムにおいて、前記対象物正規化手段は、前記対象物の位置と大きさを、前記特徴点の位置を利用して座標的に正規化する位置正規化手段と、前記特徴点の輝度をデジタル的に正規化する輝度正規化手段とからなることを特徴とする映像録画システム。

【請求項8】 請求項5に記載された映像録画システムにおいて、前記対象物照合手段は、前記対象物正規化手段による対象物の輝度信号あるいはカラー信号の正規化信号と、前記問合せ画像の輝度信号成分あるいはカラー信号成分とを比較する画素値比較手段からなり、前記対象物の輝度変化度合いが一致するか否かを所定のしきい値によって判断することを特徴とする映像録画システム。

【請求項9】 請求項5乃至8のいずれか1項に記載された映像録画システムにおいて、前記対象物は顔画像であり、前記問合せ画像指定手段は人物を指定する指定手段であり、当該顔画像は標準顔画像に対応して正規化することを特徴とする映像録画システム。

【請求項10】 入力した映像信号から所定の事物を検出した場合に録画する映像録画方法をプログラムとしてコンピュータが読み取り可能な記録媒体において、前記映像録画方法は、前記入力した映像信号の複数フレーム中から前記所定の事物を検出し、該所定の事物と指定された問合せ事物とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号を録画することを特徴とする当該映像録画方法をプログラムとするコンピュータが読み取り可能な記録媒体。

【請求項11】 入力した映像信号から所定の事物を検出した場合に録画する映像録画方法をプログラムとしてコンピュータが読み取り可能な記録媒体において、前記映像録画方法は、前記入力した映像信号の複数フレーム中から前記所定の事物を検出し、該所定の事物と指定された問合せ事物とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号の録画を行なわないことを特徴とする当該映像録画方法をプログラムとするコ

ンピュータが読み取り可能な記録媒体。

【請求項 1 2】 入力音声信号から順次所定時間のフレーズを検出し、該入力音声信号と指定された前記所定時間の問合せ音声信号とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記入力音声信号中から選択された入力音声信号を録音することを特徴とする音声録音方法。

【請求項 1 3】 入力映像信号を遅延手段で遅延し、前記入力映像信号を画像信号と音声信号とに分離し、前記画像信号をフレーム毎に区切り、該フレーム中の特徴点に応じて正規化画像信号として正規化し、前記音声信号を順次所定間隔毎に区切り、該所定時間中のレベルに応じて正規化音声信号として正規化し、予め指定された問合せ画像及び／又は前記所定間隔の問合せ音声を用意し、当該問合せ画像及び／又は問合せ音声とをそれぞれ正規化問合せ画像及び／又は正規化問合せ音声として正規化し、前記正規化画像信号と前記正規化問合せ画像及び／又は前記正規化音声信号と前記正規化問合せ音声とを照合し、当該照合に成功した場合、その時点の前後の一定の時間、前記入力映像信号中から前記予め指定された時間入力映像信号を録画する又は録画停止することを特徴とする映像録画方法。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、入力した映像信号から所定の対象事物或いは顔画像を検出した場合に、当該映像信号を選択的に録画する、あるいは、録画しない方法およびシステム及びコンピュータが読み取り可能な記録媒体に関する。

【 0 0 0 2 】

【従来の技術】

従来、テレビ放送番組の録画は、利用者がその放送番組を見ながら、録画ボタンを押下することにより行なうか、番組表等で予め番組を調べ、放送チャンネル、録画時刻の設定を行ない、録画予約機能により行なっている。いわゆるタイムシフトにより、録画後に、その録画内容をディスプレイに表示してみることが一般的である。

【0003】

このような録画という面での第1の従来技術として、特開昭61-289474号公報には、企業の機密保持等のため入門時の識別等に用いる識別装置であって、ビデオフロッピに記録された顔、指紋及び声紋に関する識別情報と被識別情報とを比較し、一致しないときには入門を拒否し、ビデオフロッピの取り出しを阻止すると共に、被識別情報をビデオフロッピに記録することにより、完全な機械化及び識別を可能にすると共に、証拠としても残すことができるようにしたことが記載されている。

【0004】

また、第2の従来技術として、特開昭63-118473号公報には、建屋又は車両などの機械式キーを用いず画像処理を応用した撮像式開施錠装置に関し、予め記憶された基準データと新たに到来する入力データとが一致したときに出入り口ドアを開く場合に、車両等の出入り口ドア付近に人の顔を画像信号に変換する撮像手段と、該画像信号から人の左右の眼の間隔、眼の縦横幅等の眼の形状及び配列関係を特徴データとして演算形成する特徴抽出手段と、予めメモリに登録した特徴データと新たな特徴データとが一致するか否かを判断し、一致する場合に開施錠指令を送出する識別判別手段とを備えたことが記載されている。

【0005】

また、第3の従来技術として、特開平3-127263号公報には、金融機関等において使用される自動取引装置に関し、ビデオカメラで撮影した顔写真と、身分を証明する書類上の顔写真とにより、本人の照合を行うようにしているので、新規の口座の開設や解約を機械により自動的に行うことができ、銀行などの金融機関の窓口業務の合理化を図ることが記載されている。

【0006】

また、第4の従来技術として、特開平6-73949号公報には、特定人のみが利用若しくは使用できる貸金庫等のような設備、装置、環境において、当該設備等の利用記録をとる利用者記録装置において、利用者の識別情報と利用者の顔写真とを撮影する2つ以上のビデオカメラを順次切り換えながら、撮像された情報を相互に関連してVISCA (Video System Control Architecture: VTRを

パソコンで制御する為の仕様) プロトコルにて動作するVTRを用いて記録し、オン・スクリーン・ディスプレイ(OSD)にて、利用日時、テープカウント情報等の付属情報を重ねて録画することが記載されている。

【0007】

また、第5の従来技術として、特開平9-35069号公報には、ビデオ信号の符号化のための被写対象物を自動検出するために、CCITT勧告H.261標準対応の低ビットレートでのビデオ符号化において、被写体人物の顔面外形及び顔面造作位置を閉曲線で特定した近似モデルを用いて自動的に検出して比較的高いビットレートで符号化し、しかも顔面及び顔面造作の画像以外の部分を比較的低いビットレートで符号化して、顔面及び顔面造作の画像を適切な符号化品質で採取送信することが可能となり、テレビ会議等の画面上での人対人の相互通信の品質及び効率が改善され、また、読唇に頼る聴覚障害者の場合のような或る種の状況において顔面の造作の良好な表現が得られるので、障害者にも有利であることが記載されている。

【0008】

【発明が解決しようとする課題】

しかしながら、上記第1乃至第5の従来技術には、セキュリティ応用を目的とする例であり、録画の応用については記載されていない。例えば、第1の従来技術では、本人の一致に顔や声を用いて、本人と一致しないときに非識別情報を出力するもので、録画するか否かなどには一切関与されていない。また、第2の従来技術には、顔の照合をして、扉の開錠・施錠をするもので、第3の従来技術には、本人の顔と、免許証の顔を照合するもので、第4の従来技術には顔の照合はないが保安目的の録画であり、第5の従来技術には特徴点の検出に一工夫を加えて特徴点を顔面の眼、口等に特定したことについて、それぞれ記載されているが、録画上の問題点には一切、着目されていない。

【0009】

また、セキュリティを目的とするのではなく、録画の応用という点では、単にタイムシフトのための録画だけでは十分ではない。

【0010】

例えば、長い報道番組の一部に、特定の出演者が出演することがわかっているとしても、正確な時刻は知らされないのが普通で、その特定の出演者が出演している場面の録画を失敗することもある。特に、録画用記録容量に余裕がない場合には、その特定の出演者が出演している場面の前後だけを録画したいとする場合に失敗する確率が高くなる。一方、また、番組の間に挿入される特定の広告番組を録画したいような場合には、録画予約することは大変困難で、可能性のある時間帯をすべて録画しなければならない。しかし、このような場合、不必要にビデオテープ等の媒体を利用し、時間的にまたは物理的な媒体の有効利用という面からいまだ無駄が多いのが現実である。さらに、特定の広告番組を提供する広告主にとって、現実には放送された時間が予定された放送時間に整合するののかどうかを知る必要がある場合に、現実のCM番組の放送を録画しておけば、その整合性が明確となり、好都合であり、この点は広告主であるクライアントと共に、放送業者にとっても、必要なことであり、このことを実現する方法及び装置が望まれていた。

【0011】

本発明の目的は、テレビ番組に出演している特定の出演者を認識し、その出演者が含まれているフレームを中心に、前後の映像を録画したり、逆に、その出演者が含まれているフレームを中心に、前後の映像を録画から削除したりする映像録画方法およびシステムを提供することにある。

【0012】

【課題を解決するための手段】

本発明による映像録画方法は、上記目的を達成するため、複数フレームの映像信号中から入力顔画像を検出し、該入力顔画像と指定された問合せ顔画像とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号を録画することを特徴とする。

【0013】

また、本発明の映像録画方法は、複数フレームの映像信号中から入力顔画像を検出し、該顔入力画像と指定された問合せ顔画像とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号の録画

を行なわないことを特徴とする。

【 0 0 1 4 】

また、本発明は、対象物を検出して前記対象物が存在するフレームの前後の複数フレームを録画する映像録画システムにおいて、映像信号から対象物の特徴点を検出する対象物検出手段と、該対象物の特徴点を利用して該対象物を正規化する対象物正規化手段と、予め問合せ画像を指定して正規化し正規化問合せ画像として出力する問合せ画像指定手段と、前記対象物正規化手段の出力と前記正規化問合せ画像とを照合する対象物照合手段と、前記対象物照合手段の結果に応じて前記映像信号を録画制御する制御手段とを備えたことを特徴とする。

【 0 0 1 5 】

また、本発明は、入力した映像信号から所定の事物を検出した場合に録画する映像録画方法をプログラムとしてコンピュータが読み取り可能な記録媒体において、前記映像録画方法は、前記入力した映像信号の複数フレーム中から前記所定の事物を検出し、該所定の事物と指定された問合せ事物とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号を録画することを特徴とする当該映像録画方法をプログラムとするコンピュータが読み取り可能な記録媒体である。

【 0 0 1 6 】

また、本発明は、入力した映像信号から所定の事物を検出した場合に録画する映像録画方法をプログラムとしてコンピュータが読み取り可能な記録媒体において、前記映像録画方法は、前記入力した映像信号の複数フレーム中から前記所定の事物を検出し、該所定の事物と指定された問合せ事物とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記複数フレーム中から前記映像信号の録画を行なわないことを特徴とする当該映像録画方法をプログラムとするコンピュータが読み取り可能な記録媒体である。

【 0 0 1 7 】

また、本発明の音声録音方法は、入力音声信号から順次所定時間のフレーズを検出し、該入力音声信号と指定された前記所定時間の問合せ音声信号とを照合し、照合に成功した場合、その時点の前後の一定の時間、前記入力音声信号中から

選択された入力音声信号を録音することを特徴とする。

【0018】

また、本発明の録画方法は、入力映像信号を遅延手段で遅延し、前記入力映像信号を画像信号と音声信号とに分離し、前記画像信号をフレーム毎に区切り、該フレーム中の特徴点に応じて正規化画像信号として正規化し、前記音声信号を順次所定間隔毎に区切り、該所定時間中のレベルに応じて正規化音声信号として正規化し、予め指定された問合せ画像及び／又は前記所定間隔の問合せ音声を用意し、当該等合わせ画像及び／又は問合せ音声とをそれぞれ正規化問合せ画像及び／又は正規化問合せ音声として正規化し、前記正規化画像信号と前記正規化問合せ画像及び／又は前記正規化音声信号と前記正規化問合せ音声とを照合し、当該照合に成功した場合、その時点の前後の一定の時間、前記入力映像信号中から前記予め指定された時間入力映像信号を録画する又は録画停止することを特徴とする。

【0019】

【発明の実施の形態】

本発明の実施形態について、図面を参照しつつ詳細に説明する。

【0020】

〔第1の実施形態〕

(1) 構成の説明

図1を参照して、本発明の代表的な実施形態による映像録画システムについて説明する。映像信号1が顔検出手段2と映像遅延手段8に入力される。この映像信号1は、通常のテレビチューナ部から出力される映像信号、あるいは、カメラ等の撮像装置から与えられるものであり、録画の際には録画装置の入力ビデオ端子に入力されるものである。ここでは、代表的に'映像信号'と書いているが、音声信号を含んでいるものでも扱いは同様である。また、映像信号1は、通常ビデオ信号として、1秒間に30フレームの各画面を連続的に組み合わせたものであり、静止画であっても、動画であっても、1画面の組み合わせとしては同様に扱える。

【0021】

図 1 において、入力された映像信号 1 は、顔検出手段 2 により人物の顔を検出し、顔正規化手段 3 で検出された顔画像を、例えば両眼の瞳間隔を所定値として正規化して、正規化画像として顔照合手段 6 に出力する。一方、人物指定手段 4 によって、顔画像データベースから希望する特定人物を、不図示のディスプレイ上に表示された画像から指定する。つぎに、指定された問合せとなる人物の問合せ顔画像を正規化して、正規化顔画像信号 5 1 として顔照合手段に入力する。顔画像照合手段 6 は、入力された正規化画像と正規化された問合せ顔画像とを比較し、一致するか否か、或いは類似度が所定値より小さく判別できない程度に近似しているか否か等を数量的に検出して、一致度合いに応じた一致信号 6 2 を制御手段 7 に出力する。制御手段 7 は、一致信号に従って、一致していれば録画する、或いは録画を停止する等の事前の指示に応じて、映像信号遅延手段 8 からの読み出し信号を映像信号録画手段 9 に録画／停止するように録画制御信号 7 1 を出力する。映像信号録画手段 9 は制御手段 7 の指示に従って、映像信号遅延手段 8 からの映像信号を録画／停止する。また、映像信号録画手段 9 は再生を指示された場合、ディスプレイ等に映像再生信号として出力する。こうして、例えば特定の有名人を撮影した映像情報を収集することが可能となる。

【 0 0 2 2 】

以下、各部分について、詳細に説明する。顔検出手段 2 は、与えられた映像信号 1 から、人物の顔を検出する。顔検出には、肌色の検出、動きの検出、顔らしい陰影部分の検出などの公知の技術が利用できる。例として、図 2 を参照して、International Conference on Image Processing (1999年10月、神戸) における発表の論文集 (Vol.1, pp.662-666) に記載されている顔検出技術に基づいて、顔検出手段 2 の構成について説明する。

【 0 0 2 3 】

図 2 によれば、映像信号 1 に対し、解像度変換手段 2 0 1 が元々の解像度から検出処理に十分な粗い解像度の低解像度画像 2 0 2 に変換する。低解像度画像 2 0 2 に対して、フィルタリング手段 2 0 3 が、Gabor wavelet 2 0 4 を作用させる。Gabor wavelet 2 0 4 は式 (1) のようなフィルタであり、各画素を中心としてコンボリューションが計算される。各 n (方向) および r (解像度) に対

するフィルタリングの出力から、目尻、目頭、唇端点などの特徴点候補 2 0 5 が得られる。

【 0 0 2 4 】

【数 1】

$$g_n^r(x) = \frac{k_r^2}{\sigma^2} \exp\left(-\frac{k_r^2 \|x\|^2}{2\sigma^2}\right) \times \left[\exp\left(i(k_n^r)^T x\right) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (1)$$

ここで、解 r と方位 n におけるウェーブレット wavelet の 2 D 空間周波数は、

$$k_n^r = k_r \begin{pmatrix} \cos(n\pi / N_{orns}) \\ \sin(n\pi / N_{orns}) \end{pmatrix}, \quad \begin{matrix} 0 \leq n < N_{orns} \\ 0 \leq r < N_{resns} \end{matrix}$$

で与えられる。

次に、特徴点候補 2 0 5 に対して、仮説検定手段 2 0 6 が顔の特徴点の位置関係を検定し、顔照合範囲信号 2 0 7 を出力する。この特徴点の検出には、特徴点の有する特徴、例えば目尻と瞳とを比較した場合には、目尻の 2 D 空間周波数の横方向は粗密の差がないが、瞳の 2 D 空間周波数の横方向は粗であるが密ではなく、目尻の 2 D 空間周波数の縦方向は粗ではないが密であり、瞳の 2 D 空間周波数の縦方向は粗であるが密ではないという相違点があり、このような特徴点の差違をテーブル化しておいて、特徴点の判定を容易にしておくことができる。

【 0 0 2 5 】

上記のように顔検出手段 2 は、顔らしい部分の位置を検出し、顔照合範囲信号 2 0 7 として出力する。

【 0 0 2 6 】

図 3 に、検出された右目の位置 2 2、左目の位置 2 3、口の位置 2 4 の例を示す。これらの位置を基準として、顔照合範囲 2 5 が決定され、上記の顔照合範囲信号 2 0 7 に対応する。つぎに、顔正規化手段 3 は、前記の映像信号 1 の顔照合範囲 2 5 の部分に一定の正規化を行ない、正規化画像 2 6 を出力する。

【 0 0 2 7 】

これを、更に図 4 の実施形態を参照して説明する。映像信号 1 から検出された顔は、一般に左右に傾いているなど、姿勢が一定していない。また、画像の輝度

やコントラストも、様々である。位置正規化手段 3 1 は、図 3 に示す右目の位置 2 2、左目の位置 2 3、口の位置 2 4 を、それぞれ右目の基準位置 3 2、左目の基準位置 3 3、口の基準位置 3 4 に写像するように、画像の回転、拡大縮小を行ない、位置正規化画像 3 5 を出力する。具体的には、例えば右目の位置 2 2、左目の位置 2 3、口の位置 2 4 の座標を、それぞれ (x_r, y_r) , (x_l, y_l) , (x_m, y_m) 、右目の基準位置 3 2、左目の基準位置 3 3、口の基準位置 3 4 の座標を、それぞれ (x_{r0}, y_{r0}) , (x_{l0}, y_{l0}) , (x_{m0}, y_{m0}) とすると、映像信号 1 の各画素 (x, y) とそれに対応する位置正規化画像 3 5 の画素 (x', y') は、口の位置 2 4 を基準とし a, b を媒介変数として、式 (2), (3) のように書くことができる。

【 0 0 2 8 】

$$x - x_m = a (x_r - x_m) + b (x_l - x_m)$$

$$y - y_m = a (y_r - y_m) + b (y_l - y_m) \dots\dots (2)$$

$$x' - x_{m0} = a (x_{r0} - x_{m0}) + b (x_{l0} - x_{m0})$$

$$y' - y_{m0} = a (y_{r0} - y_{m0}) + b (y_{l0} - y_{m0}) \dots\dots (3)$$

よって、位置正規化画像 3 5 の各画素 (x', y') について、式 (3) を a, b について解き、この a, b を式 (2) に代入して、 (x, y) を求め、映像信号 1 の画素 (x, y) の画素値を位置正規化画像 3 5 の画素値とすれば、回転、拡大縮小は完了する。

【 0 0 2 9 】

輝度正規化手段 3 6 は、位置正規化画像 3 5 の顔照合基準範囲 3 5 中の各画素の輝度を正規化し、輝度正規化画像 3 7 を出力する。具体的には、例えば位置正規化画像 3 5 の顔照合範囲 2 5 の各画素の輝度 $Y(x, y)$ (x, y は画素の位置) の平均値と標準偏差を計算し、これが輝度正規化画像 3 7 では、基準平均値と基準標準偏差になるように式 (3) により画素値を正規化する。ここで、 $Y'(x, y)$ は、輝度正規化画像 3 7 の画素値である。

【 0 0 3 0 】

【数 2】

$$Y'(x, y) = \bar{Y}_0 + \frac{\sigma_{y0}}{\sigma_y} \left(Y(x, y) - \bar{Y} \right) \quad (4)$$

この輝度正規化手段 3 6 の処理は、勿論上記のものにとられるものではなく、顔照合基準範囲 3 5 中の各画素の輝度値のヒストグラムがすべての輝度値について平坦になるようにする'ヒストグラム平坦化処理'などを利用しても構わない。

【 0 0 3 1】

図 1 の顔画像データベース 5 には、単数あるいは複数の人物の、上記のような正規化が行なわれた正規化顔画像 5 1 が予め格納されている。利用者は、録画を行ないたい番組の人物を問合せ人物として、人物指定手段 4 で指定する。人物指定手段 4 は、人物の名前をキーボードから入力する名称入力手法、顔画像データベース 5 に含まれている顔画像を表示し、マウス等のポインティングデバイスにより特定の顔画像を選択する顔画像選択手法、等多くの公知の実現形態が可能である。

【 0 0 3 2】

また、コンピュータに顔の認識技術に関して、「コンピュータによる顔の認識の研究動向」（電子情報通信学会誌 Vol.80, No.3, pp257-266: 赤松茂著作）として、人物の認識という意味での顔の認識機能について、コンピュータがこれを代替え・支援できれば、人間と機械とのインターフェースの高度化への大きな貢献が期待できるとして、顔の正面を対象として、（１）皮膚や髪の色情報、顔の対称性や輪郭を構成する曲線などのマクロな形状情報を試みる顔という対象に不変な画像特徴を求めたり、多重解像度のモザイク画像のマッチングにより coarse-to-fine な顔領域検索を行う方法と、（２）シーン中の人物像の動きに伴う動画像の時空間フィルタリングによって動く頭部を追跡し、その位置をある程度絞り込んだ後、その近傍から切り出される任意の濃淡画像の内、画像空間中に分布する顔画像集合が形成している部分空間に最も近い顔らしいシーン画像から顔領域を抽

出することにより顔の濃淡パターンによってテンプレートマッチングを行う方法と、(3) 顔と顔以外の2クラスについてモザイク化された画像パターンの標本集合から、各クラスに少数のプロトタイプを求め、入力パターンと各プロトタイプとの距離値を入力データとして2クラスを分類する多層パーセプトロンを学習によって求め、これを用いて顔領域の判定を行って、顔とその他を区別する特徴を事例の学習によって求める方法によって顔パターンを抽出することができることが記載されている。ただし、顔の正面画像から顔画像の検索を行うことを特徴としている。本実施形態においても、上述した顔パターンの検索の3つの方法のいずれを用いても良い。

【0033】

次に、顔パターンの特徴を表現する方法に、造作ベースとパターン整合との2つのアプローチがある。造作ベース (feature-based) の手法は、顔という対象の構造に関する知識を活用して目、鼻、口などの造作の形状や配置における個人性に着目し、これらの造作を表す特徴点を抽出することにより、各造作の形状や位置関係を記述するものである。一方、パターン整合 (template-based) の手法は、顔という対象に固有な輪郭や造作に関する構造的な知識を用いず、顔パターンを各画素における濃淡値の二次元配列として表現される画像としてとらえてマッチングによって識別する。後者が最近では主流となっているが、顔パターンの特徴から、特にパターン整合を目や口のような部分パターンに応用し、パターンの局所的な周期性と方向性をウェーブレット変換によって求めるGabor jetという方式が提案されている。Gabor jet方式は、顔パターンをグラフ表現し、造作ベース手法とを融合させることで、顔の認識を実現しているものである。

【0034】

つぎに、図1の顔画像データベース5からは、指定された問い合わせ人物の正規化顔画像51が顔照合手段6に出力される。この正規化顔画像51は、入力される映像信号内の顔画像とサイズの的にマッチすると共に、顔の向きや照明方向等が一致しておれば、いわゆる類似度が高くなり、照合時点での一致する度合いが高くなる。顔照合手段6は、顔正規化手段3の出力である正規化画像26と顔画像データベース5の出力である正規化顔画像51を照合する。

【 0 0 3 5 】

図 5 は、顔照合手段 6 の本実施形態による構成図である。画素値比較手段 6 1 は、正規化された画像の顔照合基準範囲 3 5 同士を、そのまま画素値により比較する。各画素における画素値の差の絶対値の和、あるいは、該画素値の差の自乗の和が閾値と比較され、閾値より小さい場合に、画素値比較手段 6 1 は一致信号 6 2 を出力する。顔照合手段 6 の実現方法としては、更に高性能な手法が公知であり、例えば上述した「コンピュータによる顔の認識の研究動向」（電子情報通信学会誌 Vol.80, No.3, pp257-266: 赤松茂著作）に記載されている造作ベース (feature-based) の手法やパターン整合 (template-based) の手法などのような手法を、ここに応用することも可能である。

【 0 0 3 6 】

図 1 の制御手段 7 は、一致信号 6 2 を受けると録画信号 7 1 を映像信号録画手段 9 に出力する。映像信号遅延手段 8 は、映像信号 1 を時間的に蓄積し、一定時間遅らせて出力するものであり、半導体メモリ、磁気ディスク、コンピュータの記憶装置とコンピュータプログラムなどで実現することができる。例えば、映像信号遅延手段 8 の遅延時間が 1 0 分である場合、映像信号録画手段 9 が映像信号遅延手段 9 から 1 0 分前の映像信号を受け取る。映像信号録画手段 9 は制御手段 7 からの録画信号 7 1 を受けている間映像信号を録画する。そのようにすると、映像信号録画手段 9 は映像信号に指定された人物が現われ、その顔が検出照合された時点の 1 0 分前からの映像を録画する。制御手段 7 が、録画信号 7 1 をどのように出力するかは、目的により設定することができる。

【 0 0 3 7 】

例えば、指定された人物が、映像信号中に検出照合されてから 1 0 分間録画信号 7 1 を出力するように設定する。この設定を変更することは、映像信号遅延手段 8 がコンピュータの記憶装置とコンピュータプログラムで実現されている場合は特に容易である。テレビ番組では、例え主役であっても、常時画面に写っているわけではない。しかし 1 0 分に一度は現われると考えられる場合には、このように設定することによって、その人物の出ている番組をほぼ間違いなく録画することが可能である。

【 0 0 3 8 】

これまでの実施形態では、本発明の利用者が指定した人物の映っている番組を録画する用途について説明してきたが、本発明では同様の構成によって、録画している番組中から利用者が指定した人物の映っている部分を削除する用途にも実現することが可能である。この場合、図 1 に示した制御手段 7 は、従来の録画のように、ある時間帯に録画を行なうように録画信号 7 1 を出力しており、指定した人物に対する一致信号 6 2 が入力された場合に、その前後の設定された時間、録画信号 7 1 の出力を停止する。

【 0 0 3 9 】

上記の説明では、指定する人物の人数については言及していないが、人物数は複数でも構わない。複数の人物 A, B, . . . を指定しておき、その内のいずれかが検出されたら録画を行い、または、いずれかが検出されたら録画停止を行なうように構成することも可能である。

【 0 0 4 0 】

本発明を有効なものとするには、顔照合が高い信頼度で成功することが必須である。それには、顔画像データベース 5 がそれぞれの人物の、多様な姿勢、多様な環境での画像を含んでおり、映像信号 1 中の顔画像がその各々と比較されることにより、信頼度が高まる。上記の実施形態では、顔画像データベース 5 にはすでに必要な人物の正規化画像 2 6 が登録されているとしている。

【 0 0 4 1 】

次に、顔画像データベース 5 を強化するシステムの実施形態を、図 6 を参照して説明する。例えば、図 1 の場合と同様に、映像信号 1 中の顔画像が顔検出手段 2 で検出され、顔正規化手段 3 によって正規化画像 2 6 とされる。これが利用者が登録したい人物である場合、利用者は人物指定手段 4 により、人物を指定して顔画像データベース 5 に登録する。

【 0 0 4 2 】

別の顔画像データベース 5 の強化方法として、次のようなものがある。多くの利用者が一般的に録画したい人物（俳優・タレント）については、その人物の了解を得、その人物の多様な姿勢・環境における顔画像データベースを作成し、配

信・配布するサービスを行なう。これにより、利用者の顔画像データベース作成の手間を省くことができる。

【 0 0 4 3 】

上記は、すべて本人の照合を顔画像で行なうことで一貫しているが、本人は必ずしも顔のみで照合するのが適当とは限らない。音声照合により本人照合を行なう技術も開発されており、これを本発明に含めることも可能である。

【 0 0 4 4 】

上記実施形態では、顔画像を照合対象としたが、顔画像以外にも、自然物、建築物など、静的な対象物、或いは動的な対象物であっても、この対象物を正規化してデータベースに格納しておけば、この対象物を正規化して照合することで、一致しておけば複数のフレーム前から、その対象物が表示画面から喪失された後まで、その対象物が表示される前後のビデオ画像を録画することができる。

【 0 0 4 5 】

また、上記実施形態では、問合せ対象物を特定することで、その人物や自然物の対象物が表出する画面の前後画像を検索することも可能であり、データベースを図書館等に格納しておき、そのデータベース中から欲しい画像の前後を取得することもでき、人手で画面を見ながら検索していた従来例とは異なり、自動的に検出することができる。

【 0 0 4 6 】

なお、本発明の実施形態では放送されているテレビ番組について、本発明が適用されるように説明しているが、すでに録画された媒体の再生信号に適用されれば、上記のような独自の映像アルバムなどの映像編集システムとすることが可能である。また、監視カメラから得られる映像信号に適用されれば、指定した人物のみの映っている部分をまとめることのできる監視システムも構成可能である。

【 0 0 4 7 】

また、放送局からの画像情報をVTRに録画している場合に、希望顔画像を検出した場合にはそのVTRに録画を続行するばかりでなく、ハードディスク等の記録媒体にその顔画像の出現する前の数フレームからその顔画像が出現しなくなる数フレームまでを録画してもよい。

【0048】

また、本発明による録画媒体としては、VTRやDVD-RAM等ばかりでなく、ハードディスクやフロッピーディスク等の種々の媒体を適用できる。

【0049】

〔第2の実施形態〕

次に、本発明の第2の実施形態として、音声による所定音声を検出した場合には、録音していたフレーズの前後のみを録音するか、或いは録音を止める例について説明する。

【0050】

図7には、音声の場合の音声照合システムのブロック図を示す。画像検出の場合には、1フレーム毎、或いは1フィールド毎に顔検出することができるが、音声の場合には、リアルタイムに、連続する音声と比較対象の音声とをパターン整合法（重ね合わせ法）で比較して、一致するか否かを判断する。特に音声の場合には、個人個人の特徴はホルマント周波数特性で判断するのが一般的であるので、この技術を利用してもよい。

【0051】

まず、音声指定手段14により、対象の音声を特定する。例えば127ms単位に有名俳優の音声を音声データベース15に含まれている音声から抽出する。音声データベース15は、この指定した127ms区間の音声のピークレベルを一定値Aとして正規化して正規化信号45をする。一方、被検出対象の音声信号11は、音声正規化手段13に入力され、127ms区間を10サンプル以上の短期間毎に区切って127ms区間を定め、その127ms区間のピークレベルを一定値Aと同一として正規化信号46を出力する。音声照合手段16では、各正規化信号45、46とをパターン整合法により、両者の差分を検出し、差分値が所定の閾値より小さい場合には、両者の正規化信号45、46とは一致すると判断する。その際のパターン整合法では、127ms区間を100サンプル程度に区分して比較する。

【0052】

このパターン整合法を実行する前に、比較する127ms区間のスペクトラム

度合いを比較してもよい。すなわち、正規化信号45の127msをスペクトラム分析して、各帯域のエネルギー量を積算しておき、このスペクトラム分析と同じ分析器に正規化信号46を通し、両者のスペクトラム分析結果と量的比較を行う。この場合、両者が一致すれば、パターン整合法によって一致するか否かを判断する。さらに詳細に類似度合いを求めるとしたら、対象音声と入力音声信号とを、個人の発声器官の物理的構造によるものと、言葉のなまりや、アクセント、話し方の癖などのために、ホルマント周波数、周波数帯域、スペクトル形状をそれぞれ比較して一致するか否かと、ホルマント周波数の時間パターンと、単語の時間長とを比較して、総合的に一致するか否かを音声照合手段で実行する。また、127msは任意に変更できるようにすれば、特徴的な音声を発生する場面を特定することもできる。

【0053】

音声照合手段16では、一致したときに一致信号12を出力し、制御手段17で録音するか否かの制御を行う。この制御手段17は、音声信号遅延手段18にも指示を出し、一致信号12を入力したならば、その前後の所定時間、例えば1分程度の録音を音声信号録音手段19に録音する。或いは一致する音声の前後の録音を阻止することとしてもよい。

【0054】

上述したように、図7を用いて、顔照合の代わりに音声照合により録音を制御するシステムの実施形態を説明したが、さらに、顔照合と音声照合を組み合わせで録画・録音を制御するシステムについて、図1と図7とを参照して説明する。

【0055】

通常、映像信号は画像信号と音声信号とを含む場合、映像信号1内の画像信号は、例えばNTSC方式の場合、周波数帯域幅を0～4.5MHzとし搬送波を1.25MHzとして、音声信号は画像信号の搬送波周波数より4.5MHz高く設定して、周波数変調としてある。従って、図1の映像信号1の入力に対して、図7の音声信号入力には、映像信号1から高域成分を通す音声分離回路を設けて音声信号11を得る。この場合、音声分離回路には、FM復調回路を備えて、画像信号と分離して、音声搬送波信号を周波数変調波をベースバンド信号に復調

して、音声信号11として入力する。また、映像信号遅延手段8と音声信号遅延手段とは共通としてよく、所定時間遅延する。この遅延手段としては、ハードディスクでもよいし、CD-R、CD-RWであってもよい。

【0056】

つぎに、画像データベース5と音声データベース15とは共通でもよく、別個でもよいし、人物指定手段4と音声指定手段14も共通でよく、キーボードやマウスを用いて、問合せ対象を指定する。この場合、指定するのは、音声の所定区間、或いは画像信号の1フレーム中の顔画像或いは対象事物を指定する。この指定は、音声信号のみ、或いは画像信号のみを別個に、論理積として指定してよく、又は音声信号と画像信号とが重複する論理和として指定してもよく、この指定手段は不図示のディスプレイ上に表示して希望の問合せ対象物又は問合せ対象音を特定する。

【0057】

また、顔正規化手段3で顔画像や事物画像を正規化し、データベースからも問合せ対象物を正規化した正規化画像として顔照合手段6又は事物照合手段にて照合し、一致信号が得られた場合には、或いは類似点が多く所定のしきい値より小さくなった場合にはほぼ一致しているとして一致信号を制御手段に出力する。音声信号においても、同様に、音声照合手段16によって、問合せ音声と入力された音声信号とを照合して、一致している場合の一致信号或いは類似度合いが高い場合の一致信号を制御手段17に出力する。

【0058】

次に、制御手段7は、映像信号録画手段9に、一致信号を検出されたフレームの前後の録画を録画制御信号71として指示し、制御手段17は一致信号を検出された時点の前後の所定時間の録音を録音制御信号18として指示する。この指示は、指定手段4、14で指定した内容によって、柔軟に対応した録画・録音が可能である。

【0059】

この録画・録音組み合わせシステムでは、問合せする対象が音声だけの場合でも、入力映像信号中から一致した場合には録音・録画も可能なので、例えばテレ

ビ等に頻繁に出演する有名人の情報を入手したいとした場合、特定の顔画像を指定し、その有名人の音声をも指定しておけば、その有名人の画像は勿論、音声だけが入力された場合でも、録音・録画が可能なので、その有名人の情報を隈無く収集できるという相乗的な効果を発揮できる。

【 0 0 6 0 】

【発明の効果】

以上に説明したように、本発明を利用することにより、利用者が指定した人物の出演している部分のみを録画することができ、不要な録画部分を節約することができる効果がある。また、同様に利用者が指定した人物の出演している部分のみを録画から除くことも可能である。

【 0 0 6 1 】

また、通常の時刻を指定することによる録画方法では、前の番組が延長されたような場合には、所望の番組を適切に録画することができないが、本発明を利用すれば、自動的に録画時刻が変更されるため、所望の番組の録画が可能である。

【 0 0 6 2 】

本発明は、ビデオテープレコーダ、ビデオディスクレコーダ等に組み込むことによって、これらの商品に新たな機能を付加することが可能である。また、コンピュータとその記憶装置を利用して実現すれば、独自の映像アルバムのような映像の編集等の自動化にも利用することができる。

【図面の簡単な説明】

【図 1】

本発明の典型的な実施形態である映像録画システムの構成の説明図である。

【図 2】

本発明に用いる顔検出手段の構成例の説明図である。

【図 3】

本発明による顔検出手段により検出された顔の特徴点および顔正規化手段により正規化された顔の特徴点の説明図である。

【図 4】

本発明に用いる顔正規化手段の実施形態の説明図である。

【図 5】

本発明に用いる顔照合手段 6 の単純な実施形態の説明図である。

【図 6】

顔画像データベースを強化するシステムの実施形態の説明図である。

【図 7】

本発明の典型的な実施形態である音声録音システムの構成の説明図である。

【符号の説明】

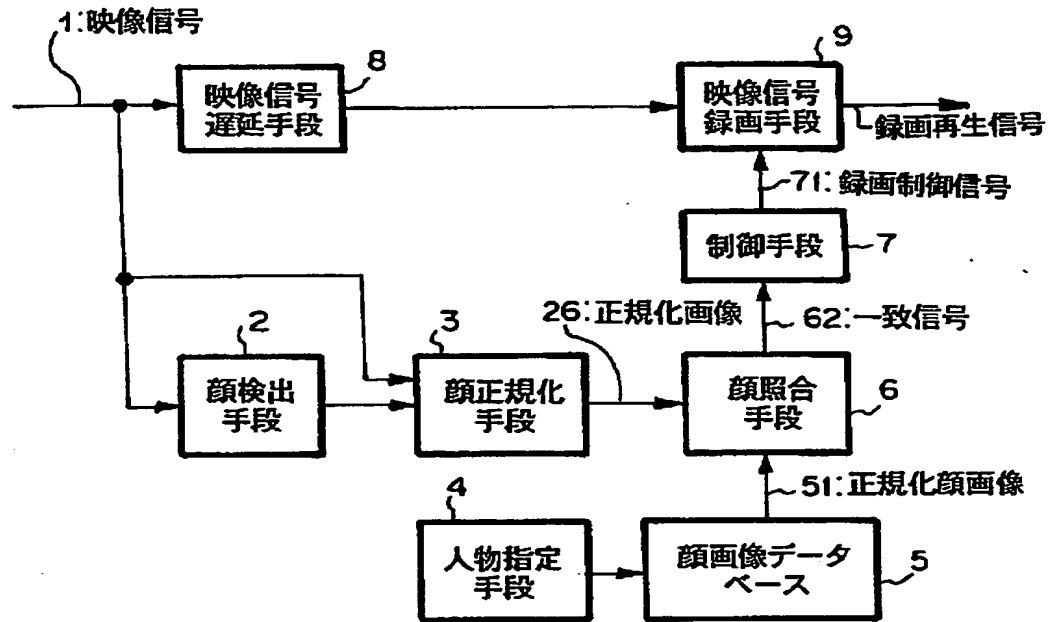
- 1 映像信号
- 2 顔検出手段
- 3 顔正規化手段
- 4 人物指定手段
- 5 顔画像データベース
- 6 顔照合手段
- 7 制御手段
- 8 映像信号遅延手段
- 9 映像信号録画手段
- 2 2 右目の位置
- 2 3 左目の位置
- 2 4 口の位置
- 2 5 顔照合範囲
- 3 1 位置正規化手段
- 3 2 右目の基準位置
- 3 3 左目の基準位置
- 3 4 口の基準位置
- 3 5 顔照合基準範囲
- 3 6 輝度正規化手段
- 6 1 画素値比較手段
- 2 0 1 解像度変換手段
- 2 0 3 フィルタリング手段

特 2 0 0 0 - 0 9 9 6 2 7

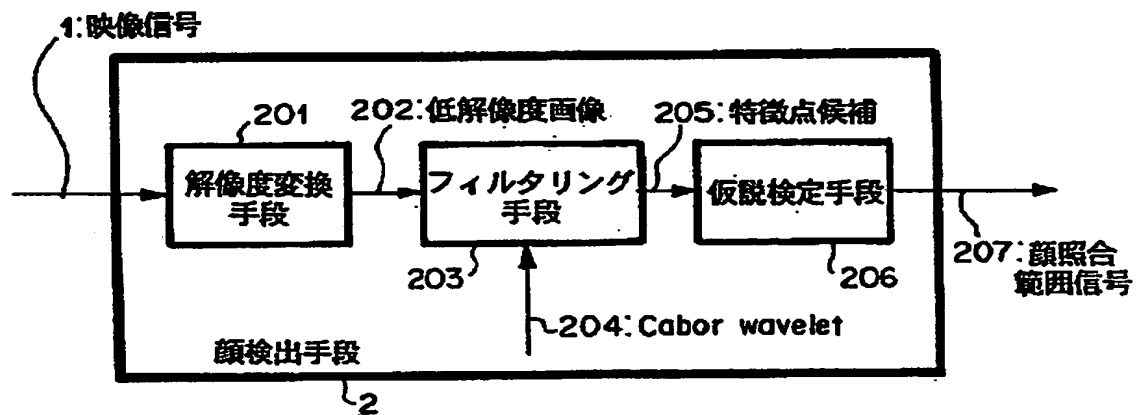
2 0 6 仮説検定手段

【書類名】 図面

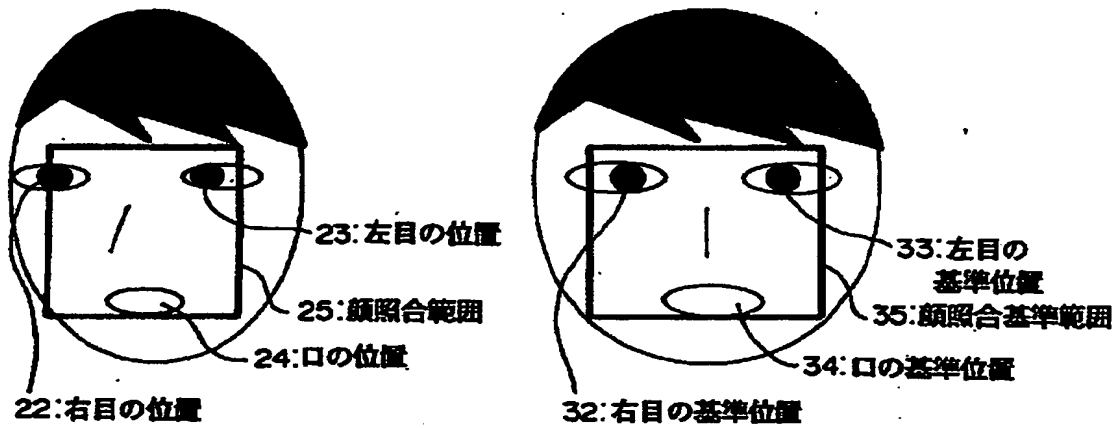
【図 1】



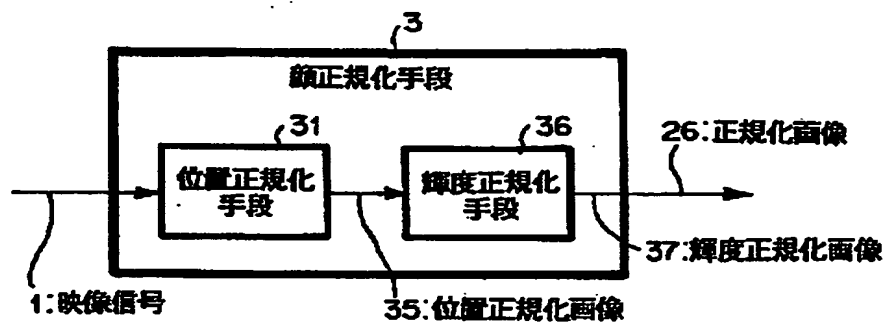
【図 2】



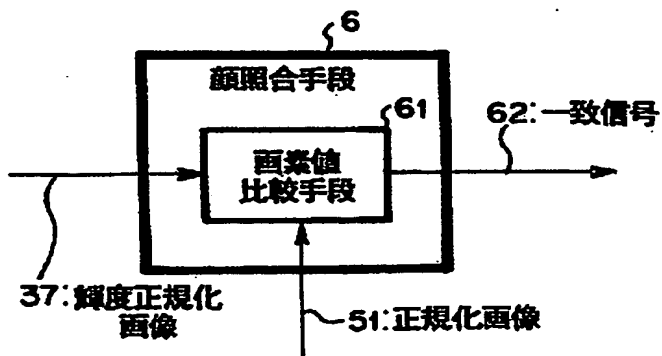
【図 3】



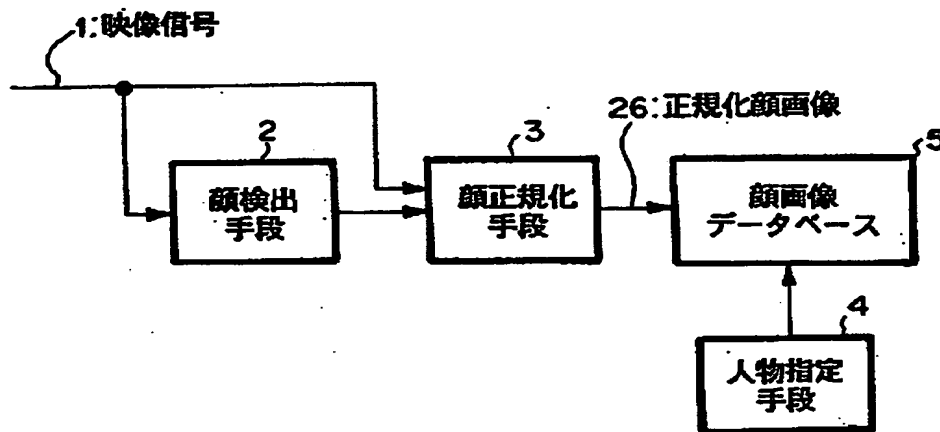
【図 4】



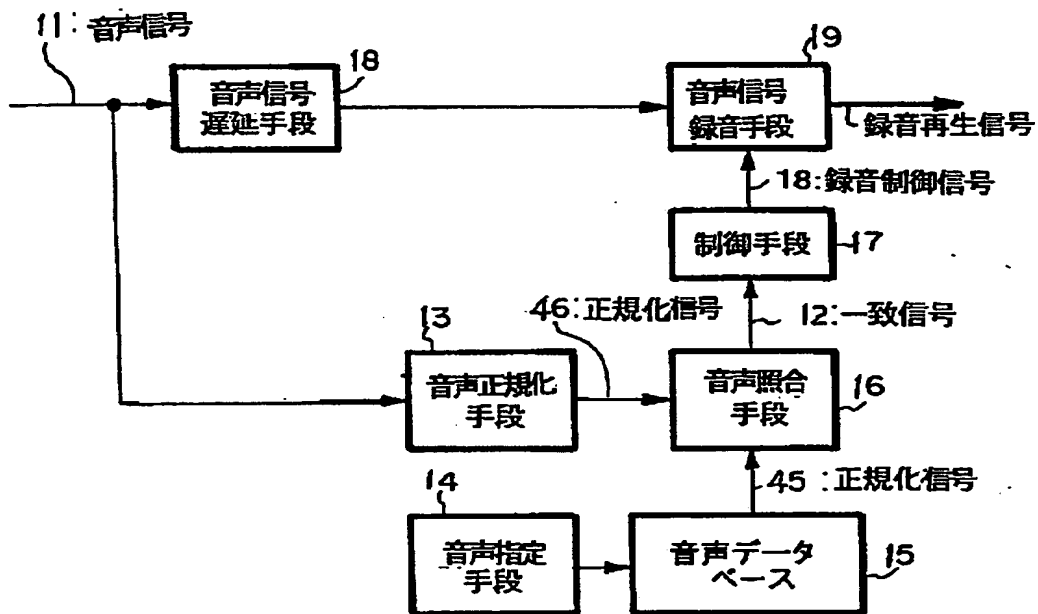
【図 5】



【図 6】



【図 7】



【書類名】 要約書

【要約】

【課題】 番組内容に関係なく開始・終了時刻を設定した場合、詳細な番組内容が知らされていない場合、あるいは、番組の予定が変更になった場合に、利用者の所望の番組を画像内容に応じて、録画する方法・システムを提供することを課題とする。

【解決手段】 映像信号中の出演人物の顔を自動的にデータベースと照合し、所望の人物が出演している前後のフレームを録画する、あるいは、録画しないように制御することにより、録画を行なう。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000004237]

1. 変更年月日	1990年 8月29日
[変更理由]	新規登録
住 所	東京都港区芝五丁目7番1号
氏 名	日本電気株式会社